

NYCHVS Microdata Documentation for ASA Data Challenge Expo

Each survey cycle of the NYCHVS has microdata available in 3 files from the Census: household-level occupied, household-level vacant, and person-level occupied.

Custom-Formatted Datasets

To make these data more accessible for participants in the ASA Data Challenge Expo, we have created a set of microdata comprising all ten survey cycles from 1991-2017. This set contains only household-level occupied records. For the convenience of participants, these set of ten microdata files are available in three formats, each containing identical data:

1. Stata DTA
2. SAS XPORT
3. Comma-delimited CSV

We have created variable names that are consistent across all ten survey cycle files for each variable/item in these files only.

Other NYCHVS Microdata

The original ASCII flat files can be downloaded from the Census website (<https://www.census.gov/programs-surveys/nychvs/data/datasets.All.html>). Person-level and/or household-level vacant files may be downloaded and combined with either the household-level occupied files available on census.gov or to the custom datasets provided here.

Please note that the files available on census.gov do not contain variable names; if referencing the original flat files, users must specify them on their own.¹ For these files available from Census, please reference questionnaires and record layouts/codebooks available on their website.

Variable Names and Variable Crosswalk

We have created a Variable Crosswalk by Survey Cycle to assist users in identifying available variables across survey cycles and locating them in their codebooks. Please reference the Variable Crosswalk by Survey Cycle to see the item number for each question in each year and how they can be found in a survey cycle's corresponding codebook. On the HPD website, we have provided the 2017 NYCHVS Questionnaire and codebooks for each of the provided files. Please note that though we have aligned consistent variable names for the same items across survey cycles, values may differ from one survey to the next. As such, users are advised to reference the relevant survey cycle codebook(s) as appropriate.

¹ Note that 2014 and 2017 come with import code on the website. Variable names in these programs are specific to the Census and do not match the files provided here and have not been aligned with previous survey cycles.

The variable names assigned in these files generally follow the item numbers in the codebooks. For example, householder's sex is variable "_1f," which corresponds to item 1f on the questionnaire. For variables with no item number in the codebook(s) (generally, these are recoded items), the variable name takes a general abbreviated form of the item name. For example, "Monthly Gross Rent" is assigned the variable "mgrent." There are exceptions: as mentioned, some variable names do not match item numbers to allow them to be consistent across the 10 survey cycles of NYCHVS files. Thus, some variables do *not* conform to their item number as it is listed in the record layout or questionnaire for that year. For ease of identification, these variables whose aligned variable name do not match their item numbers in a given survey cycle have been shaded gray in the Variable Crosswalk.

Linking Multiple Survey Cycles

The nature of the NYCHVS is that it is longitudinal across survey cycles within a decade. Within each decade, users can use the variable called "linkage" to link records across survey cycles.²

Weights

Users should also note that the NYCHVS is designed such that the records represent the overall New York City housing stock at the time of survey. The survey cycles from 1991-2014 have single household weights "hhweight" for each record. The data from 2017 have replicate weights; however, the household weight "hhweight" can be used the same as in previous survey cycles.

² Please note that files can only be used longitudinally within a given decade (i.e., 1991, 1993, 1996 and 1999; 2002, 2005, and 2008). 2011, 2014, and 2017 are longitudinal in nature but cannot be linked because of data disclosure avoidance protections.